α-carboxysome size is controlled by the disordered scaffold protein CsoS2

Luke M. Oltrogge*[,a,b], Allen W. Chen*[,c], Thawatchai Chaijarasphong[†,a], Julia B. Turnšek[a], David F. Savage[‡,a,b,d]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, California, USA

[b]Howard Hughes Medical Institute, University of California, Berkeley, California, USA

[c]Department of Chemistry, University of California, Berkeley, CA, USA

[d]Innovative Genomics Institute, University of California, Berkeley, California, USA

*These authors contributed equally to this work.

[†]Present address: Department of Biotechnology, Faculty of Science, Mahidol University, Rama 16 VI Rd., Bangkok 10400, Thailand

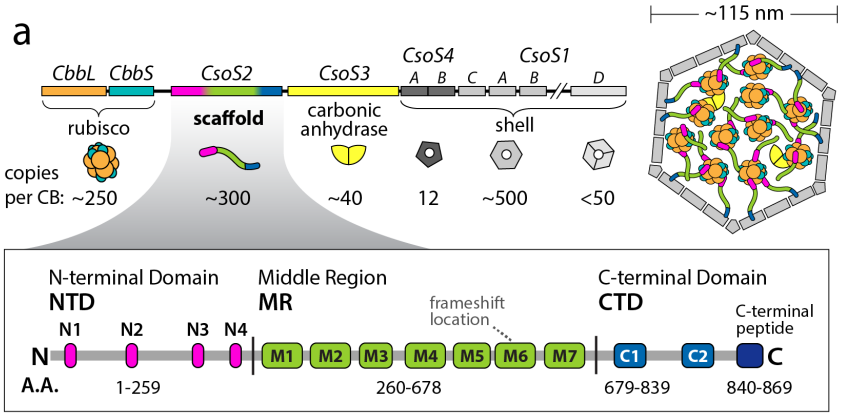
[‡]To whom correspondence should be addressed: savage@berkeley.edu

Abstract:

Carboxysomes are protein microcompartments that function in the bacterial CO_2 concentrating mechanism (CCM) to facilitate CO_2 assimilation. To do so, carboxysomes assemble from thousands of constituent proteins into an icosahedral shell which encapsulates the enzymes rubisco and carbonic anhydrase to form structures typically >100 nm and >300 megadaltons. Although many of the protein interactions driving the assembly process have been determined, it remains unknown how size and composition are precisely controlled. Here we show that the size of α-carboxysomes is controlled by the disordered scaffolding protein CsoS2. CsoS2 contains two classes of related peptide repeats which bind to the shell in a distinct fashion, and our data indicate that size is controlled by the relative number of these interactions. We propose an energetic and structural model wherein the two repeat classes bind at the junction of shell hexamers but differ in their preferences for the shell contact angles, and thus the local curvature. Thus, in contrast to other models driven by cargo or shell alone, our data show that CsoS2 and shell proteins interact to collectively achieve the large size and monodispersity of α-carboxysomes.

Introduction:

Spatial organization is a fundamental feature of biology–delineating the boundaries between cells and the environment and segregating the internal metabolism.[1] Cyanobacteria and diverse chemoautotrophic bacteria use specialized protein microcompartments called carboxysomes to provide a CO_2 enriched micro-environment for the carboxylase rubisco which fixes CO_2 into sugars.[2–4] Carboxysomes are a striking example of emergent order, with thousands of individual copies of ten different constituent proteins spontaneously assembling into quasi-icosahedral particles with well-defined size and composition.[5] The mechanism of size control is particularly interesting, as the assembled particles extend far beyond the dimensions of any individual component. The assembly process also requires high-fidelity with potential for severe phenotypic penalty. Morphological defects to the carboxysome, such as by deletion of shell genes, break the CO_2 concentrating mechanism and result in cell death.[6,7]



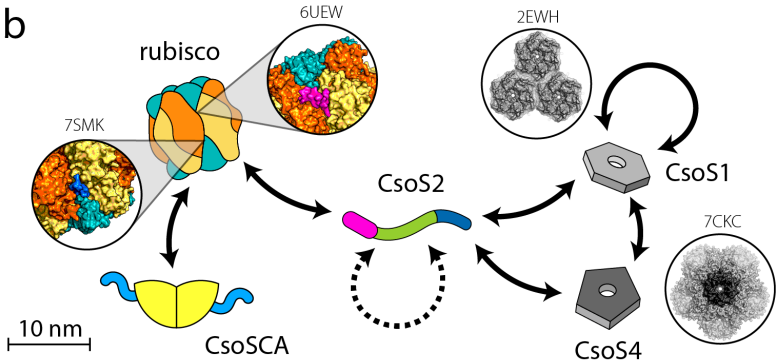


Figure 1
a) The *H. neapolitanus* regulon with the ten proteins making up the α-carboxysome. The callout shows the domain structure with repeated peptide motifs in the disordered scaffold protein CsoS2. Note that M7 has been reclassified as an M-peptide (see Fig. S1). **b**) Carboxysome protein interactome. Known protein-protein interactions are depicted with solid arrows and cases where the contact structures are solved are shown in the popouts along with their corresponding PDB IDs. CsoS2 is an important hub in the network bridging the cargo to shell connection. Cartoon schematics are shown roughly to scale.

Carboxysomes occur in two evolutionarily distinct lineages: α, found in most marine cyanobacteria and several clades of bacterial chemoautotrophs; and β, found in certain freshwater cyanobacteria. β-carboxysomes tend to be larger (100-400 nm in diameter) and are thought to assemble in an inside-out manner with the formation of a dense rubisco kernel preceding encapsulation by the shell proteins. [8 9 10] α-carboxysomes are somewhat smaller (80-120 nm in diameter) and more regular in size. Their assembly is hypothesized to involve the concurrent formation of cargo and shells. [11 12,13] In this work we focus on α-carboxysomes to understand what factors drive the size and regularity of the particles.

α-carboxysome assembly culminates in particles that are densely packed with cargo, free of shell defects to serve as a CO_2 permeability barrier, and of uniform size with diameters of about 115 nm in the model species *Halothiobacillus neapolitanus*. [4,14] The process by which carboxysome size is determined is not understood, though several models have been put forward. [15 16 17] Structural studies on the shell proteins have revealed spontaneous assembly of small icosahedral complexes with triangulation numbers of T = 3, 4, or 9.[18 19 20] By comparison, the *H. neapolitanus* carboxysome is significantly larger with T ~ 75. [21] This suggests that intrinsic curvature preferences of the shell proteins alone[§] are not determinative of the carboxysome size–as is the case for many small icosahedral virus capsids–but that an interplay with internal components is ultimately responsible for particle size.

α-carboxysomes are composed of a thin icosahedral shell made up of thousands of hexameric protein capsomers (CsoS1) and pentameric capsomers (CsoS4) forming the twelve vertices.[21] Inside the shell is a dense cargo of enzymes: rubisco and carbonic anhydrase, and the scaffolding protein CsoS2 (Fig. 1). [22–25] CsoS2 is unique among the carboxysome proteins as it is predicted to be largely structurally disordered along its entire length. [4,26 12] It does, however, contain several distinct repeated sequence motifs that collectively define a three-part domain structure: the N-terminal domain (NTD) with N-peptide motifs, the middle region (MR) with M-peptide motifs, and the C-terminal domain (CTD) with C-peptide motifs and a conserved C-terminal peptide (CTP). Previous work has shown that CsoS2 is essential to growth and acts as a central node with interactions to both rubisco and CsoS1.[12,13 27] Specifically, the NTD binds to and encapsulates rubisco via interactions with the N-peptides, [13,24] and the CTD has been shown to associate with the shell.[19,20] We recently showed that MR also interacts with the shell, though what distinguishes its functional role from the CTD is not yet clear. [28]

Here we describe how the sequence of the scaffold CsoS2 and, in particular, the relative numbers of M- and C-peptide motifs specify the size of the carboxysome. Furthermore, this size effect is largely independent of the primary cargo rubisco and points toward a model for size control that is driven by the shell's mechanical properties and, specifically, CsoS2's modulation of these mechanical properties. Finally, in light of a recent cryoEM structure of portions of the CTD-shell interface [20] we propose a structural model of the M-peptide's role.

[§] These studies with shell-only complexes use a subset of capsomer paralogs. It is unknown whether other combinations might change size characteristics.

Materials and Methods:

Cloning and expression

All recombinant carboxysome expression was based on the pHnCB10 plasmid which contains the full carboxysome regulon from *Halothiobacillus neapolitanus* composed of ten proteins under the control of a lac operator, has a ColE1 origin, and carries a chloramphenicol resistance marker. [29] For testing different CsoS2 variants and truncations, a Golden Gate destination vector was created from pHnCB10, replacing the original CsoS2 with lacZ bracketed by BsaI type II restriction sites. The different CsoS2 constructs were built by PCR amplification of the desired fragments with primers containing flanking BsaI sites and compatible recombination sites. The polyproline II helix sequence was purchased as a gBlock from IDT. All plasmids were completed by Golden Gate Assembly. [30] The co-expression plasmids for CsoS1ABC (Fig. 2b), CsoS2A, CsoS2B (Fig. 3c), and fragments of the MR fused to GFP (Fig. S4) were built on a pFA backbone with a p15a origin, kanamycin resistance, and tetracycline operator. [31] All constructs are summarized in Table S1.

Plasmids were transformed into electrocompetent BW25113 *E. coli* cells. For carboxysome expression, cells were grown at 37°C to mid-log (OD600 0.4-0.6) whereupon the temperature was reduced to 18°C and inducer was added: 1 mM IPTG for all samples and a variable amount of anhydrotetracycline (aTc) for the variable co-expression. Cultures were grown overnight then pelleted and frozen for later use or taken straight away for carboxysome purification.

H. neapolitanus cells were grown at 30°C in DSMZ 68 media. The construction of the strain expressing carboxysomes with only CsoS2B from a genomic neutral site under control of a lac operator is described in ref. [32]. To express carboxysomes and sustain atmospheric growth, this strain requires IPTG which we included at 1mM for this purpose.

Carboxysome purification

Carboxysomes were purified as described previously. [27] The cells from 1L expression cultures were lysed with 25 mL B-PER bacterial lysis reagent (Thermo Fisher) supplemented with 0.01 mg/mL DNaseI, 0.1 mg/mL lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM $MgCl_2$, and 20 mM $NaHCO_3$. Lysate was clarified by centrifugation at 12,000 rcf for 30 min. Carboxysomes were pelleted from the supernatant by centrifugation at 40,000 rcf for 30 min. Raw pelleted carboxysomes were resuspended with 20 mL TEMB buffer (10 mM Tris, 10 mM $MgCl_2$, 1 mM EDTA, and 20 mM $NaHCO_3$, pH 8.4) on ice and pelleted again with a 40,000 rcf spin for 30 min. The pellet was again gently resuspended in 2mL TEMB and loaded on top of sucrose gradient with 5 mL layers of 10, 20, 30, 40, and 50% sucrose (w/v) in TEMB. The gradient samples were spun in an ultracentrifuge for 35 min. at 105,000 rcf. Then the gradients were fractionated into 1 mL fractions and analyzed by SDS-PAGE to determine the carboxysome-containing fractions which also correspond to bluish light scattering from the particles. Fractions with carboxysomes were pooled, centrifuged for 90 min. at 105,000 rcf, resuspended with TEMB, and stored at 4°C prior to negative stain transmission electron microscopy (TEM).

143 *Transmission electron microscopy and size characterization*

144 Carboxysome samples were diluted to a 280 nm absorbance of 0.05-0.1. Prior to
145 sample application Formvar/carbon coated, copper EM grids were prepared by glow
146 discharged. 5 µL of sample was loaded to each grid, allowed to sit for at least 2 min.,
147 washed and stained with continuous application of 10 uL of 1% (w/v) uranyl acetate
148 solution, and dried with filter paper. In our experience it is important to minimize the
149 exposure of the carboxysomes on the grid to air-water interfaces because these can
150 promote carboxysome breakage on the surface and thus precluding accurate size
151 analysis. All samples were imaged either with a FEI Tecnai 12 120 kV or JEOL 1200 EX
152 80 kV transmission electron microscope.

153 Particle size characterization was performed in ImageJ. Carboxysomes were
154 manually located to avoid particle misidentification and exclude obviously damaged
155 carboxysomes. The images were first given a pixel offset of 4 counts to eliminate any
156 zero count pixels. Then polyhedra were drawn around each carboxysome and the pixel
157 counts zeroed within this region. Finally, the particles were selected with a zero count
158 threshold and measured with the automatic particle characterization utility. The
159 minimum Feret diameter–i.e. the closest approach of two parallel lines in contact with
160 the particle projection–was used for all size statistics.

161 *MR localization to carboxysomes*

162 A series of plasmids of MR truncations fused to Superfolder GFP [33] were each
163 co-transformed with the full carboxysome plasmid. Carboxysome expression was
164 induced with 1 mM IPTG and GFP fusion expression was induced with 100 nM aTc in
165 mid-log cells. Temperature was reduced to 18°C and the cells were grown overnight.
166 Carboxysomes were purified as described above. Then for each sample the
167 absorbance was measured at 340 nm and GFP fluorescence measured with 485 nm
168 excitation and 515 nm emission on a Tecan M1000 plate reader. The carboxysomes
169 strongly scatter at 340 nm while the protein has low absorbance at this wavelength.
170 Thus the GFP fluorescence intensity divided by the 340 nm absorbance was used as a
171 proxy for the relative amount of GFP loading per carboxysome.

172

173 *Modified CsoS2 in* H. neapolitanus

174 *H. neapolitanus* is naturally competent and will uptake foreign DNA and integrate
175 it into its genome by homologous recombination given suitable flanking homology arms.
176 We used this to knock out CsoS2 at the native locus with a spectinomycin resistance
177 cassette. This intermediate strain is incapable of atmospheric growth due to its inability
178 to form carboxysomes. We then knocked in a series CsoS2 variants having variable M-
179 peptide repeat numbers into the genomic neutral site NS1 along with a kanamycin
180 resistance gene. All variants could grow in CO_2-replete conditions with 5% CO_2. Colony
181 forming unit (CFU) counts were performed by serial dilution of liquid cultures grown at
182 5% CO_2 and plated on DSMZ 68 agar plates with spectinomycin (20 µg/mL), kanamycin
183 (10 µg/mL), and IPTG (1 mM) and placed in the corresponding CO_2 environments:
184 atmosphere (~0.04%), 0.5%, and 5%.

Structural modeling of M-peptide shell interaction

The AlphaFold predicted structures of four widely divergent CsoS2 proteins were downloaded from Uniprot (RefIDs): [34] *Halothiobacillus neapolitanus* (O85041), *Prochlorococcus marinus* (Q7V2C8), *Nitrosomonas eutropha* (Q0AHV9), and *Imhoffiella purpurea* (W9VGM3). The coordinates matching the M-peptides were extracted and compared pairwise for structural similarity using TMAlign. [35] They were commonly aligned to M4 from *N. eutropha* as the nearest structure to the centroid.

H. neapolitanus M1-peptide was docked to the luminal side of the junction of three CsoS1A hexamers. The coordinates of this shell trimer were obtained from symmetry expansion of the coordinates of PDB ID 2EWH. [36] Rosetta was used to independently relax the structures of the M1-peptide and the shell trimer and the best scoring structures were used for protein-protein docking. The M1-peptide was positioned above the three-fold shell axis and roughly aligned with the M1-peptide pseudo-threefold axis. Next the Rosetta local docking routine was used with initial translation randomization of 5Å and orientational randomization of 12°. 500 docked structures were generated. [37] The best scoring among these clearly fell into a common class of configurations with the VTGs wedged near the symmetry related CsoS1A-His79 residues. The protein-protein interface was relaxed for the structure shown in Fig. 6c.

Bioinformatic characterization of M- and C-peptide repeats

All sequences from IMG with the CsoS2M pfam (PF12288) were downloaded from IMG May 2020 for a total of 770 sequences. These were filtered to eliminate those with partial sequence and unspecified residues. Next the sequences were dereplicated to 95% identity using usearch for a total of 272 unique sequences. [38] MEME was used to identify peptide motifs and finding, as previously reported, the main known motifs of the N-peptide, M-peptide, C-peptide, and CTP. [39] MAST from the MEME suite was used to extract the positions of all matches to the M- and C-peptide motifs and the amino acid sequences were extracted for each of these with a 15aa buffer zone on either side. These peptide sequences were aligned all together (ie. both M- and C-peptides) using mafft. [40] FastTree was used to generate a phylogenetic tree on the basis of this alignment. [41] From the tree there are clearly distinguishable clades for the M- and C-peptide classes but some of the initial MAST hits were mislabeled (including *H. neapolitanus* M7). All sequences belonging to each clade were aligned against each other and Weblogo3 was used to generate sequence logos (Fig. 6a, Fig. S1). [42]

Results:

Size in self-assembling biological systems can be controlled by a number of mechanisms including intrinsic curvature built into the geometry of the constituent proteins, [43] molecular rulers, [44] excluded volume effects, [17] and non-equilibrium race conditions between growth and termination. [15] We hypothesized whether rubisco, which is approximately 65% of the particle by mass[5] and most of the cargo, might play a pivotal role in guiding the carboxysome size and overall morphology. However, several studies have shown that expression of carboxysome particles without rubisco yields the same basic icosahedral architecture and size, including even in the native *Halothiobacillus* host. [19 45 46 47] We repeated this experiment – by recombinantly expressing all carboxysome proteins except rubisco in our *E. coli* heterologous expression system – and obtained similar results (Fig. 2a). In particular, the particle size does not collapse to that of the much smaller complexes formed from shell proteins alone (Fig. 2a, blue lines).

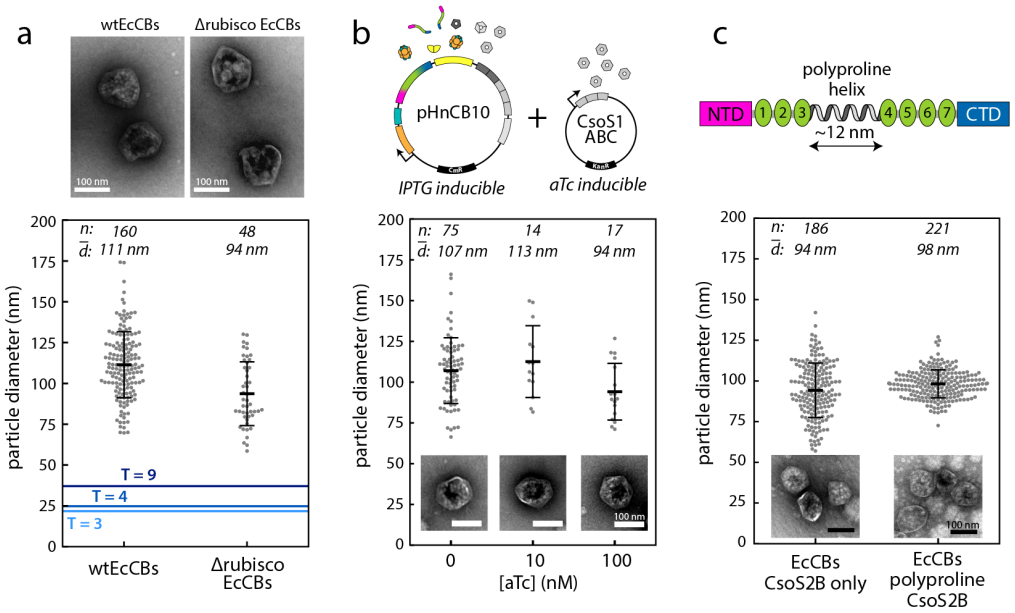


Figure 2

a) Comparison of recombinant carboxysomes with and without the primary cargo rubisco. TEM negative stain micrographs highlight the same overall pseudo-icosahedral structure. All swarm plots show the minimum Feret diameter. The dark central line is mean diameter with whiskers to +/- 1 standard deviation. Particle counts and mean diameters are shown along the top. The blue lines mark the sizes of solved icosahedral complexes of shell capsomers and the corresponding T-numbers are indicated. [19,20] **b**) Hexamer titration. *E. coli* carrying pHnCB10, a plasmid containing all ten carboxysome proteins under control of lac operator, and a plasmid with the major shell proteins CsoS1ABC under tetracycline induction control. Cells were all induced with 1 mM ITPG and variable amounts of anhydrotetracycline (aTc) to change shell expression levels. Swarm plots of particle sizes as in a) with inset zoomed TEM micrographs showing normal morphology. **c**) Polyproline helix. Carboxysomes were expressed with CsoS2B, ie. CsoS2 with no frameshift, and CsoS2B with a 12 nm polyproline II helix inserted. Swarm plots of particle sizes as in a) with inset TEM micrographs of representative particles.

We next turned to the major shell components: the hexameric proteins. One hypothesis is that increasing shell concentration would alter the balance between cargo and shell growth in favor of the shell to enclose smaller particles, ultimately short-circuiting the assembly of full-sized particles. [15] To test this idea, the three major hexameric proteins CsoS1ABC were co-expressed with variable induction alongside the full set of carboxysome proteins. The size and morphologies of the resulting particles were not appreciably different; however, the presence of extra shell proteins was deleterious to the yield of purified carboxysomes (Fig 2b), likely by consuming other carboxysome components into off-pathway assemblies and/or aggregates.

A third hypothesis is that CsoS2, also a major component at ~15% by mass, [5] is a possible governor of carboxysome size. To this point, it has previously been found that disordered scaffold proteins can play a role in the size specification of certain icosahedral virus capsids. [48,49] A particularly striking case is PBCV-1 in which a disordered protein acts as a molecular ruler, effectively measuring the distance between vertices.[44] We thus tested whether CsoS2 acts as a molecular ruler by inserting a sequence encoding a polyproline II helix – a rigid structural element – of approximately 12 nm between the M3 and M4 repeats.[27,50] This construct also produced carboxysomes, but this structural modification of CsoS2 showed no discernable size effect (Fig. 2c).

Given these negative results, we postulated whether other known biochemistry within the carboxysome related to size control. In *H. neapolitanus* carboxysomes CsoS2 appears as two distinct isoforms: a short form of ~61 kDa called CsoS2A and a long form of ~92 kDa called CsoS2B. We previously showed that the short form is produced as a result of a -1 programmed ribosomal frameshift (PRF) that leads to premature termination (Fig. 3a). [27] In native *H. neapolitanus* carboxysomes and ones produced recombinantly in *E. coli*, CsoS2A and CsoS2B occur in approximately equimolar quantities. Furthermore, CsoS2B alone can still sustain atmospheric growth in *H. neapolitanus* while CsoS2A cannot and does not assemble purifiable carboxysomes. [27,32] We thus asked whether this balance of scaffold isoforms has consequences for particle size. Carboxysomes from *H. neapolitanus* were purified from the wild-type strain (having both CsoS2A and B) and a mutant strain in which CsoS2 was knocked out at the native locus and complemented with a CsoS2B-only sequence–lacking the frameshifting slippery sequence–into a neutral site. Similarly, carboxysomes were expressed and purified from *E. coli* with either the wild-type (wtCsoS2) sequence or the CsoS2B-only sequence in pHnCB10. There was good agreement between the average diameters of carboxysomes obtained from *H. neapolitanus* and *E. coli*. However, a reduction of size was observed in both going from wtCsoS2 to CsoS2B carboxysomes with mean diameters changing from 114 nm to 93 nm for *H. neapolitanus* and 113 nm to 93 nm for *E. coli* (Fig. 3b), similar to the effect of eliminating rubisco (Fig. 2a). While appearing fairly modest, this represents a nearly 40% decrease of volume without the frameshifting element.

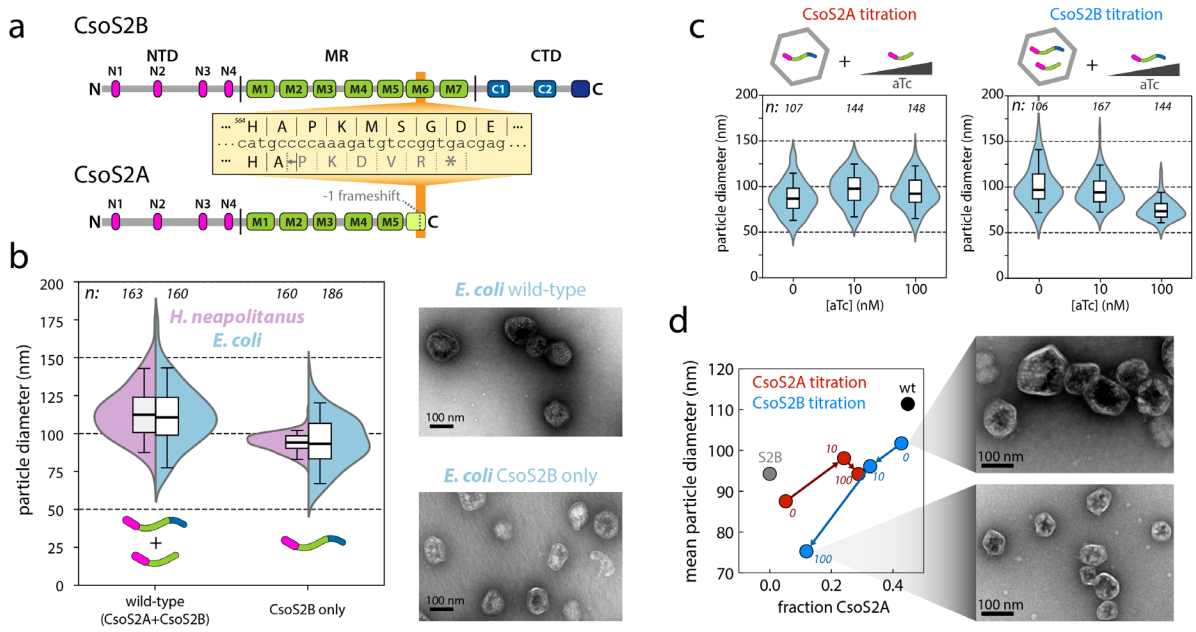
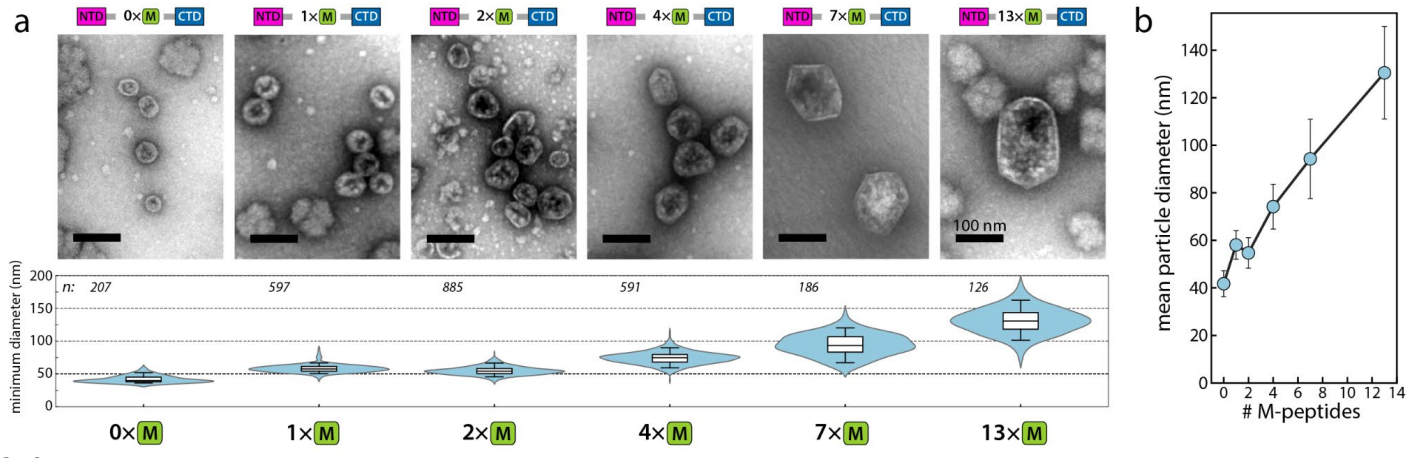


Figure 3

a) Diagram of the programmed ribosomal frameshift (PRF) leading to ribosome slippage to -1 frame and premature termination of CsoS2 to generate the CsoS2A isoform. **b**) Kernel density estimate distributions of carboxysome sizes from *H. neapolitanus* and *E. coli* expressing carboxysomes with wtCsoS2 (CsoS2A + CsoS2B) or only CsoS2B in which the PRF has been eliminated. White boxes span the inner quartiles with the median as the central black line. Whiskers are from the 5th to the 95th percentile. Representative negative stain TEM images are shown for the *E. coli* samples. **c**) Scheme and distributions for titration of CsoS2A on a background of CsoS2B-only carboxysomes (red) and titration of CsoS2B on a background of wtCsoS2 carboxysomes. **d**) Mean particle diameters for titration conditions plotted versus the measured fraction of CsoS2 as CsoS2A from gel densitometry. Popouts are representative TEM micrographs of the extremes of the CsoS2B titration showing substantially smaller carboxysomes with overexpression of CsoS2B.

To explore this connection further, we employed a titratable co-expression system in *E. coli.* In the first, CsoS2A was titrated against a background of carboxysomes with CsoS2B only. In the second, CsoS2B was titrated against a background of wild-type carboxysomes, having both CsoS2A and CsoS2B (Fig. 3cd). Broadly, the results indicate that the balance of CsoS2A and CsoS2B systematically alters the size of the resulting carboxysomes. That is, more CsoS2A correlates with larger particle sizes. Interestingly, overexpression of CsoS2B enhanced the recovered yield of carboxysomes while also reducing the mean particle sizes even lower than that of carboxysomes having only CsoS2B (Fig. 3b, Fig. 3d). This suggests that whatever interaction is driving the size effect may not be fully saturated under normal assembly conditions.



321 ***Figure 4***
322 **a**) A series of carboxysomes were produced with variable numbers of M-peptide repeats going
323 from 0 to 13 (CsoS2B has 7). Negative stain TEM images are shown for representative particles
324 in the population. All images are at the same scale with 100 nm scale bars. Below is the
325 distribution of minimum particle Feret diameters. Central black line indicates the median, white
326 box represents inner quartiles, and whiskers extend to the 5th to 95th percentile. **b**) Plot of
327 mean particle diameters as a function of the number of M-peptides. Whiskers are +/- 1 standard
328 deviation of distribution.

329

330 CsoS2A and CsoS2B are distinguished by the absence or presence of the CTD,
331 respectively, which we know to be an indispensable feature for carboxysome assembly.
332 [27] Wild-type and CsoS2B-only carboxysomes then differ in the relative amount of MR
333 with respect to the CTD. To isolate this effect, we therefore built a series of CsoS2
334 variants in which we systematically truncated or augmented the MR to include different
335 numbers of the conserved M-peptide repeats. In all of these, we eliminated the PRF
336 sequence such that resulting proteins would be precisely defined with respect to the M-
337 peptide to CTD ratio without the complicating effects of frameshifting.

338 Each CsoS2 variant produced purifiable carboxysomes that encapsulated
339 rubisco. Across the series of M-peptide numbers we observed a dramatic size trend
340 (Fig. 4). The smallest carboxysomes were those with no M-peptide repeats and with an
341 average diameter of just over 40 nm, a size which is highly similar to the single particle
342 structure solved by Ni et al. [20] at T=9 and 36.9 nm diameter. In contrast, augmenting
343 CsoS2B with copies of M-peptides M1-M6 resulted in particles having substantially
344 larger diameters, with an average of 130 nm versus 94 nm for unmodified CsoS2B.

345 In addition, we tested a subset of these variants for complementation in the
346 wtCsoS2 knockout *H. neapolitanus* strain. We found that the MR truncations negatively
347 impacted survival at atmospheric CO_2 concentrations (~0.04%) with 1 and 4 M-peptide
348 repeats unable to grow while 7 and 13 M-peptide repeats could. At 0.5% CO_2 all but the
349 1 M-peptide strain could survive. All strains could grow under CO_2 replete conditions
350 (5%) (Fig. S2).

The two C-peptides of the CTD are related to the M-peptides and share a common motif of three VTG triplets each separated by 8 residues (Fig. S1). This similarity raises an obvious question – do these C-peptides play a similar role as the M-peptides with respect to carboxysome size? To test this hypothesis, we created two CsoS2 variants: one with a duplication of both C-peptide repeats and another with duplication of the entire CTD, including the CTP. In contrast to the MR augmentation which produced substantially larger carboxysomes, the C-peptide duplications did not result in larger particles and, instead, produced similarly sized or even slightly smaller carboxysomes compared to CsoS2B (Fig. S3).

It is now well established that the CTD forms interactions with the carboxysome shell via both the C-peptides and CTP. [19,20] Our recent study found that the MR also specifically interacts with the shell hexamers and that the VTG triplets are an essential component of that interaction. [28] To further test this interaction in the context of carboxysome assembly we designed a series of MR-GFP fusion proteins having different numbers of M-peptides. These were co-expressed alongside complete carboxysomes, and we quantified the relative GFP loading (Fig. S4). We found a strong dependence on the M-peptide repeat number, with more repeats resulting in more loading. This result implies that the MR interacts with the shell during assembly and does so in a multivalent fashion.

In summary, the M-peptides and C-peptides both share similar sequence features and interact with the shell. However, they have divergent effects on the size: the number of M-peptides strongly dictates the particle size while the C-peptides are absolutely required for carboxysome-like particle formation, but their relative numbers seemingly do not alter particle size.

Discussion:

In this study we have examined the factors determining the size of α-carboxysomes. We discovered that the particle size is strongly specified by the M-peptide repeat number, and that this size preference is transmitted via its interaction with the shell. Below, we discuss the implications of these observations and put forward a size control model with a potential structural mechanism.

The dispensability of rubisco to the formation of carboxysome-like particles implies that it is not an active participant in the assembly itself. Instead it acts more as a client protein, partitioning into a phase-separated condensate with CsoS2 via the NTD on the surface of the growing shell. [13] Thus we are left with the interplay between CsoS2 and shell from which to build a model. This model must take into account several observations: 1) the CTD and MR both interact with shell proteins; [19,20,28] 2) their interactions are both multivalent in nature and due to the repeated M- and C-peptides, respectively [20,28] (Fig. S4), 3) the M- and C-peptides share a sequence motif – the VTG triplet – so that they likely have related modes of contacting the shell (Fig. S1); and 4) the M- and C-peptides differ in their effects on particle size, i.e. M-peptides encourage larger particles while C-peptides are size neutral (Fig. 4, Fig. S3). Taken together, we believe that these factors point towards a model in which CsoS2 modifies the mechanical properties of the shell.

It has been previously shown that expressing shell proteins alone results in small icosahedral structures, revealing that, in the absence of other factors, the capsomers form stable high-curvature contacts. [18,51 52] Introduction of CsoS2 into this mix frustrates the formation of these small particles and instead produces significantly larger particles with lower curvature at the shell-shell interfaces. We specifically propose that the essence of the size control lies in energetic differences of shell binding between the M- and C-peptides with respect to the angles at the junctions of shell hexamers. This energetic effect can be expressed as two related but distinct models we call “preferred curvature” and “flexible curvature” (Fig. 5a). In preferred curvature, the C-peptide has an energy minimum at a greater shell contact angle than the M-peptide. In the flexible curvature model the C-peptide and M-peptide have the same energy minimum angle, however, the C-peptide has a broader energy function and can tolerate greater contact angles than the M-peptide. Note that the angle, while represented as a two-dimensional plane angle in the cartoon schematic (Fig. 5a), is likely a three-dimensional solid angle. Finally, it should be emphasized that the model is not for C-peptide binding at shell junctions directly bordering the pentameric vertex capsomer – where Ni et al. find the CTP [20] – but rather for stabilizing hexamer-only junctions in the vicinity of the vertex that must sustain higher curvature than those further away for which the M-peptide is favored (Fig. 5b). [53]

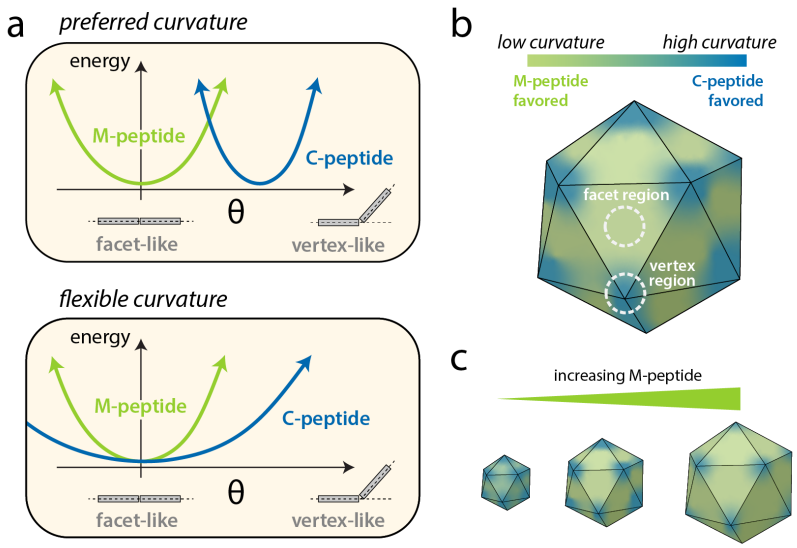


Figure 5
a) Energetic models of M- and C-peptide shell interactions. Plots show potential energy of shell-bound M- (green) and C-peptides (blue) as a function of shell angle. In the "preferred curvature" model the C-peptide has an energy minimum for a high curvature shell configuration. In the "flexible curvature" model the C-peptide has the same preferred angle as the M-peptide but a shallower energy well. **b**) Idealized map of carboxysome shell curvature. Higher curvature regions are centered on vertices while lower curvature prevails in between and on facets. The model predicts green regions will be more favorable for M-peptide binding and blue regions will be more favorable for C-peptide binding. **c**) More M-peptide stabilizes particles with a greater share of low curvature surface and thus greater diameter.

This model can be used to interpret a number of experimental observations. First, the M-peptide series in Fig. 4 entails increasing M-peptide numbers on a constant background of two C-peptides. To a first approximation, the number of high curvature sites will be roughly the same as particles increase in size while the number of low curvature sites will scale with the surface area. Thus, smaller particles are energetically better suited to a low M- to C-peptide ratio, while larger particles will benefit from a high M- to C-peptide ratio (Fig. 5c). [53] Second, the insufficiency of CsoS2A alone to form carboxysomes may be due to the fact that, while it could perhaps knit together flat sheets of hexamer facets with M-peptides, it does not possess the C-peptides necessary to stabilize the regions of highest curvature near the vertices and is thus incapable of producing closed particles. Finally, the larger size of *H. neapolitanus* carboxysomes with wtCsoS2 is explained by the extra abundance of the M-peptides relative to CsoS2B-only to extend the low curvature between vertices.

With this phenomenological model in hand can we make any inferences into its structural basis? At first glance this seems improbable given CsoS2's high disorder score throughout the sequence. [12,13,27] However, analysis of a handful of highly diverse CsoS2 sequence structural predictions as part of the AlphaFold Uniprot database revealed that, while the overall structures are amorphous and low confidence, the

aligned M-peptide sequences give a remarkably consistent microdomain (Fig. 6a). This structure had the VTG triplet motifs arranged into a triangular structure with pseudo-threefold symmetry. Furthermore, the conserved pair of cysteine residues were located in close proximity to one another. This is notable because the carboxysome interior is thought to possess an oxidized environment, and thus one might expect a fold-stabilizing disulfide bridge to form. [8,54,55]

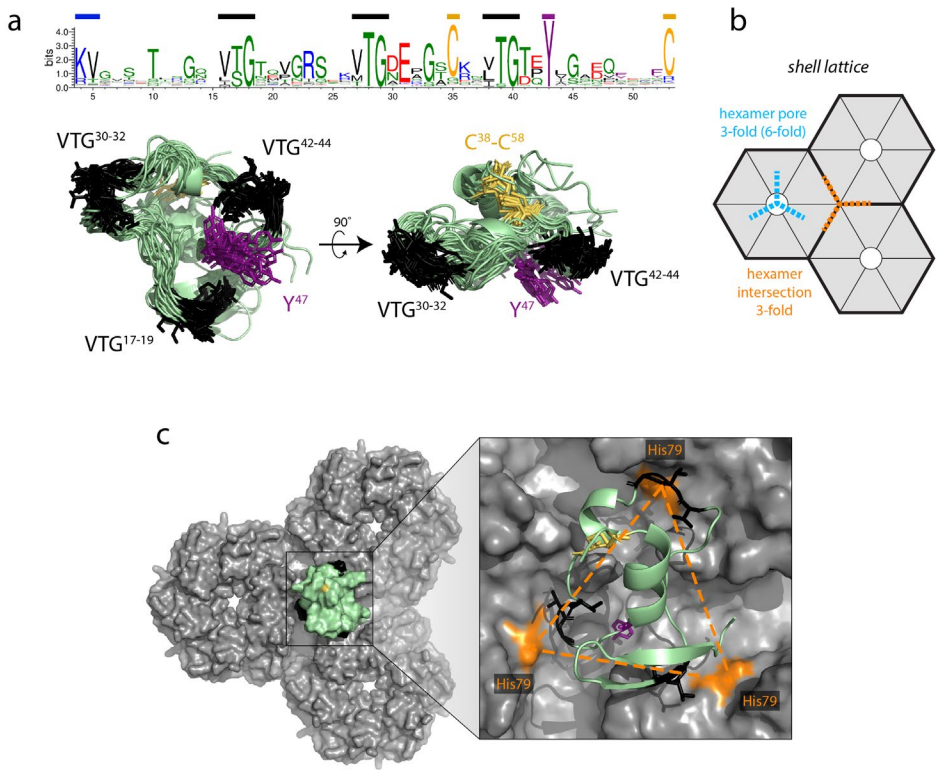


Figure 6
a) Sequence logo for M-peptide created from 1662 M-peptide sequences pulled from 272 CsoS2 sequences after dereplication to 95% identity. Aligned AlphaFold predictions for all M-peptides in CsoS2s from *Halothiobacillus neapolitanus, Prochlorococcus marinus, Nitrosomonas eutropha,* and *Imhoffiella purpurea*. Key conserved residues are colored matching the bars on the sequence logo. **b**) Schematic of hexameric shell lattice indicating the site classes with three-fold symmetry. **c**) Surface representation of *H. neapolitanus* M1-peptide docked onto the three-fold hexamer junction with Rosetta. Zoom in shows that the VTG motifs come close to matching the triangle defined by CsoS1A-His79.

The above structural prediction combined with the curvature model suggests that this pseudo-threefold symmetric M-peptide ought to bind at a shell junction with three-fold symmetry (Fig. 6b, orange). We took the AlphaFold prediction for the *H. neapolitanus* CsoS2 M1-peptide and used Rosetta to dock the structure and relax the interface onto the three-fold hexamer junction using the CsoS1A coordinates (PDB ID:

467 2EWH). [36,37] The fit has good size and shape complementarity and seemingly
468 represents a plausible structure with some additional conformational adaptation. A
469 cryoEM structure with portions of the CTD resolved was very recently described by Ni et
470 al. [20] Most interestingly, they identified CsoS1A-His79 as an important contact for CTD
471 binding through the VTG motifs. His79 is also in close proximity to the points of the
472 predicted VTG triangle (Fig. 6c).

473 What might structurally distinguish the C-peptide from the M-peptide and explain
474 their energetic preferences for different shell curvatures? The M-peptide consensus
475 sequence is essentially the C-peptide consensus plus a few additional conserved
476 positions, namely the aforementioned cysteine pair, a tyrosine residue appearing three
477 residues downstream of the last VTG motif, and a lysine-valine pair near the beginning
478 (Fig. 6a, Fig. S1). In our study of the MR we observed significant effects on the
479 biochemical interactions with the shell and sufficiency of resulting carboxysomes for
480 atmospheric growth in *H. neapolitanus* for the conserved tyrosine, highlighting a distinct
481 yet crucial role of the M-peptides. [28] The AlphaFold predictions for the C-peptides,
482 unlike the M-peptides, are of low confidence and lack a common structure. The CTD-
483 shell structure from Ni et al. shows that even as the C-peptide and CTP residues are
484 well resolved, particularly at the VTG binding contacts, the overall structure threads its
485 way along the hexamers without a clear tertiary structure of its own. [20] We hypothesize,
486 in contrast, that the M-peptide acts more as a microdomain with a characteristic fold
487 (Fig. 6a). All together, we think this suggests that the M-peptide motif is more rigid and
488 can only bind to low angle hexamer junctions while the C-peptide motif is more flexible
489 and can accommodate higher angle junctions.

490 Given this model of size we can make some inferences about the natural
491 diversity of CsoS2 sequences and the attending size consequences. Broadly speaking,
492 we predict that organisms with high M- to C-peptide ratios will produce on average
493 larger carboxysomes while those with low M- to C-peptide ratios will produce smaller
494 carboxysomes. Furthermore, we also predict that, all else being equal, CsoS2
495 sequences containing a PRF will afford larger carboxysomes than an otherwise identical
496 sequence. It will be interesting to see whether these predictions map onto
497 environmental conditions. For example, we observed that CsoS2 truncated to 4 M-
498 peptides could not sustain growth in *H. neapolitanus* under atmospheric CO_2
499 concentrations but could at 0.5% CO_2. Perhaps organisms with small predicted
500 carboxysomes occupy environments with elevated CO_2.

501 More broadly, how does this mechanism of size control compare to other
502 bacterial microcompartments (BMCs)? β-carboxysomes do not form large icosahedral
503 shell structures in the absence of rubisco [51] nor do they possess an obvious analogue to
504 CsoS2. Combined with the evidence for inside-out assembly, this suggests that the size
505 is determined as a result of a close interplay between a growing shell stabilized around
506 a dense cargo. [8,15,16] Among other BMCs (e.g. Pdu and Eut) there is a wide variety of
507 sizes and morphologies. It is not presently clear whether they fall under common
508 assembly principles and size controls. Interestingly, some BMC shell proteins can
509 assemble into diverse macrostructures such as nanotubes, extended sheets, and
510 rosettes.[56] A leading hypothesis for the BMC size control is that it may be driven by the
511 relative stoichiometries of constituent shell capsomers that differ in their toleration of

curvature. [18,53,57,58] With additional structural and biochemical data, it will thus be interesting to see whether there is a unifying principle for size control of microcompartments broadly or if, as with viruses, there are multiple mechanistic routes to achieve similar outcomes.

Conclusions:

In this study we sought to uncover the determinants of particle size in α-carboxysomes. We discovered that the size is set primarily by the scaffold protein CsoS2 and its interactions with the shell. Through systematic modification of CsoS2 we observed that the size effect is mediated by the relative numbers of M-peptide repeats found in the MR and the C-peptide repeats found in the CTD. Additionally, we determined that a functional consequence of the PRF in *H. neapolitanus* CsoS2 is to increase particle size at no cost of additional sequence.

We propose an energetic model based on differential curvature preferences between the M- and C-peptides and connect this model to a mechanistic structural hypothesis on the basis of AlphaFold structural predictions and new experimental breakthroughs in carboxysome structure.

An increasingly comprehensive view of α-carboxysome assembly is coming into focus as the structures and interactions linking the components together are elucidated. Only a few key gaps – most notably the CsoS2-MR shell interaction – remain before a complete account of the structure, spanning atomic detail to functional particles. With this depth of information we think that the carboxysome is a useful paradigm for complex self-assembly and potentially a template, physically but also conceptually, for future engineering efforts.

Acknowledgements:

We acknowledge the staff at the UC Berkeley Electron Microscope Laboratory for training and assistance with TEM. This work was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, & Biosciences Division, Physical Biosciences Program, under Award Number DE-SC0016240 to D.F.S. and D.F.S. is an Investigator of the Howard Hughes Medical Institute.

Competing interests:

D.F.S. is a co-founder of Scribe Therapeutics and a scientific advisory board member of Scribe Therapeutics and Mammoth Biosciences. All other authors declare no competing interests.

549 **References:**
550
551 (1) Lee, H., DeLoache, W. C., and Dueber, J. E. (2012) Spatial organization of enzymes
552 for metabolic engineering. *Metab. Eng. 14*, 242–251.
553 (2) Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH
554 determines the energetic efficiency of the cyanobacterial CO2 concentrating
555 mechanism. *Proc Natl Acad Sci USA 113*, E5354-62.
556 (3) Espie, G. S., and Kimber, M. S. (2011) Carboxysomes: cyanobacterial RubisCO
557 comes in small packages. *Photosyn. Res. 109*, 7–20.
558 (4) Heinhorst, S., Cannon, G. C., and Shively, J. M. (2006) Carboxysomes and
559 Carboxysome-like Inclusions, in *Complex intracellular structures in prokaryotes*
560 (Shively, J. M., Ed.), pp 141–165. Springer Berlin Heidelberg, Berlin, Heidelberg.
561 (5) Sun, Y., Harman, V. M., Johnson, J. R., Brownridge, P. J., Chen, T., Dykes, G. F.,
562 Lin, Y., Beynon, R. J., and Liu, L.-N. (2022) Decoding the Absolute Stoichiometric
563 Composition and Structural Plasticity of α-Carboxysomes. *MBio 13*, e0362921.
564 (6) Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S.
565 (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome
566 shell to function as a CO2 leakage barrier. *PLoS ONE 4*, e7521.
567 (7) Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M., and Cannon, G.
568 C. (2008) CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes
569 lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J.*
570 *Biol. Chem. 283*, 10377–10384.
571 (8) Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A., and Polka, J. K.
572 (2013) The bacterial carbon-fixing organelle is formed by shell envelopment of
573 preassembled cargo. *PLoS ONE 8*, e76127.
574 (9) Cameron, J. C., Wilson, S. C., Bernstein, S. L., and Kerfeld, C. A. (2013) Biogenesis
575 of a bacterial organelle: the carboxysome assembly pathway. *Cell 155*, 1131–1140.
576 (10) Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., Long, B. M.,
577 Price, G. D., Hartl, F. U., and Hayer-Hartl, M. (2019) Rubisco condensate formation by
578 CcmM in β-carboxysome biogenesis. *Nature 566*, 131–135.
579 (11) Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013) Functions,
580 compositions, and evolution of the two types of carboxysomes: polyhedral
581 microcompartments that facilitate CO2 fixation in cyanobacteria and some
582 proteobacteria. *Microbiol. Mol. Biol. Rev. 77*, 357–379.
583 (12) Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S.,
584 Shively, J., Cannon, G. C., and Kerfeld, C. A. (2015) Advances in understanding
585 carboxysome assembly in prochlorococcus and synechococcus implicate csos2 as a
586 critical component. *Life (Basel) 5*, 1141–1171.
587 (13) Oltrogge, L. M., Chaijarasphong, T., Chen, A. W., Bolin, E. R., Marqusee, S., and
588 Savage, D. F. (2020) Multivalent interactions between CsoS2 and Rubisco mediate α-

589 carboxysome formation. *Nat. Struct. Mol. Biol. 27*, 281–287.
590 (14) Shively, J. M., Ball, F., Brown, D. H., and Saunders, R. E. (1973) Functional
591 organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus
592 neapolitanus. *Science 182*, 584–586.
593 (15) Rotskoff, G. M., and Geissler, P. L. (2018) Robust nonequilibrium pathways to
594 microcompartment assembly. *Proc Natl Acad Sci USA 115*, 6341–6346.
595 (16) Mohajerani, F., and Hagan, M. F. (2018) The role of the encapsulated cargo in
596 microcompartment assembly. *PLoS Comput. Biol. 14*, e1006351.
597 (17) Mohajerani, F., Sayer, E., Neil, C., Inlow, K., and Hagan, M. F. (2021) Mechanisms
598 of Scaffold-Mediated Microcompartment Assembly and Size Control. *ACS Nano 15*,
599 4197–4212.
600 (18) Sutter, M., Greber, B., Aussignargues, C., and Kerfeld, C. A. (2017) Assembly
601 principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science 356*,
602 1293–1297.
603 (19) Tan, Y. Q., Ali, S., Xue, B., Teo, W. Z., Ling, L. H., Go, M. K., Lv, H., Robinson, R.
604 C., Narita, A., and Yew, W. S. (2021) Structure of a Minimal α-Carboxysome-Derived
605 Shell and Its Utility in Enzyme Stabilization. *Biomacromolecules 22*, 4095–4109.
606 (20) Ni, T., Jiang, Q., Ng, P. C., Shen, J., Dou, H., Zhu, Y., Radecke, J., Dykes, G. F.,
607 Huang, F., Liu, L., and Zhang, P. (2023) Intrinsically disordered CsoS2 acts as a
608 general molecular thread for α-carboxysome shell assembly. *bioRxiv*. DOI:
609 10.1101/2023.06.24.546370v1
610 (21) Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C.,
611 and Yeates, T. O. (2008) Atomic-level models of the bacterial carboxysome shell.
612 *Science 319*, 1083–1086.
613 (22) Blikstad, C., Dugan, E. J., Laughlin, T. G., Liu, M. D., Shoemaker, S. R., Remis, J.
614 P., and Savage, D. F. (2021) Discovery of a carbonic anhydrase-Rubisco supercomplex
615 within the alpha-carboxysome. *BioRxiv*. DOI: 10.1101/2021.11.05.467472v1
616 (23) Metskas, L. A., Ortega, D., Oltrogge, L. M., Blikstad, C., Lovejoy, D. R., Laughlin, T.
617 G., Savage, D. F., and Jensen, G. J. (2022) Rubisco forms a lattice inside alpha-
618 carboxysomes. *Nat. Commun. 13*, 4863.
619 (24) Ni, T., Sun, Y., Burn, W., Al-Hazeem, M. M. J., Zhu, Y., Yu, X., Liu, L.-N., and
620 Zhang, P. (2022) Structure and assembly of cargo Rubisco in two native α-
621 carboxysomes. *Nat. Commun. 13*, 4299.
622 (25) Kerfeld, C. A., and Melnicki, M. R. (2016) Assembly, function and evolution of
623 cyanobacterial carboxysomes. *Curr. Opin. Plant Biol. 31*, 66–75.
624 (26) Baker, S. H., Lorbach, S. C., Rodriguez-Buey, M., Williams, D. S., Aldrich, H. C.,
625 and Shively, J. M. (1999) The correlation of the gene csoS2 of the carboxysome operon
626 with two polypeptides of the carboxysome in thiobacillus neapolitanus. *Arch. Microbiol.*
627 *172*, 233–239.
628 (27) Chaijarasphong, T., Nichols, R. J., Kortright, K. E., Nixon, C. F., Teng, P. K.,

629 Oltrogge, L. M., and Savage, D. F. (2016) Programmed Ribosomal Frameshifting
630 Mediates Expression of the α-Carboxysome. *J. Mol. Biol. 428*, 153–164.
631 (28) Turnšek, J. B., Oltrogge, L. M., and Savage, D. F. (2023) Repetitive and conserved
632 motifs in an intrinsically disordered protein drive α-carboxysome assembly., *manuscript*
633 *in preparation.*
634 (29) Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A.,
635 and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle. *Proc Natl*
636 *Acad Sci USA 109*, 478–483.
637 (30) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision
638 cloning method with high throughput capability. *PLoS ONE 3*, e3647.
639 (31) Flamholz, A. I., Dugan, E., Blikstad, C., Gleizer, S., Ben-Nissan, R., Amram, S.,
640 Antonovsky, N., Ravishankar, S., Noor, E., Bar-Even, A., Milo, R., and Savage, D. F.
641 (2020) Functional reconstitution of a bacterial CO2 concentrating mechanism in
642 Escherichia coli. *eLife 9*.
643 (32) Chaijarasphong, T. (2016, October) Towards an In Vitro Reconstitution of the α-
644 Carboxysome. Doctoral dissertation, UC Berkeley.
645 (33) Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S.
646 (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat.*
647 *Biotechnol. 24*, 79–88.
648 (34) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
649 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C.,
650 Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,
651 Adler, J., Back, T., and Hassabis, D. (2021) Highly accurate protein structure prediction
652 with AlphaFold. *Nature 596*, 583–589.
653 (35) Zhang, Y., and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm
654 based on the TM-score. *Nucleic Acids Res. 33*, 2302–2309.
655 (36) Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S.,
656 Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein
657 shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol. 5*, e144.
658 (37) Marze, N. A., Roy Burman, S. S., Sheffler, W., and Gray, J. J. (2018) Efficient
659 flexible backbone protein-protein docking for challenging targets. *Bioinformatics 34*,
660 3461–3469.
661 (38) Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST.
662 *Bioinformatics 26*, 2460–2461.
663 (39) Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015) The MEME Suite.
664 *Nucleic Acids Res. 43*, W39-49.
665 (40) Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment
666 software version 7: improvements in performance and usability. *Mol. Biol. Evol. 30*,
667 772–780.
668 (41) Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2 — approximately

669 maximum-likelihood trees for large alignments. *PLoS ONE 5*, e9490.
670 (42) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a
671 sequence logo generator. *Genome Res. 14*, 1188–1190.
672 (43) Mannige, R. V., and Brooks, C. L. (2009) Geometric considerations in virus capsid
673 size specificity, auxiliary requirements, and buckling. *Proc Natl Acad Sci USA 106*,
674 8531–8536.
675 (44) Fang, Q., Zhu, D., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., Chen, Z., Sun, Y.,
676 Gross, M. L., Van Etten, J. L., Zhang, X., and Rossmann, M. G. (2019) Near-atomic
677 structure of a giant virus. *Nat. Commun. 10*, 388.
678 (45) Menon, B. B., Dou, Z., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008)
679 Halothiobacillus neapolitanus carboxysomes sequester heterologous and chimeric
680 RubisCO species. *PLoS ONE 3*, e3570.
681 (46) Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., Dykes, G. F., He,
682 H.-L., Wang, Q., Sprick, R. S., Cooper, A. I., and Liu, L.-N. (2020) Reprogramming
683 bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.*
684 *11*, 5448.
685 (47) Jiang, Q., Li, T., Yang, J., Aitchison, C. M., Huang, J., Chen, Y., Huang, F., Wang,
686 Q., Cooper, A. I., and Liu, L.-N. (2023) Synthetic engineering of a new biocatalyst
687 encapsulating [NiFe]-hydrogenases for enhanced hydrogen production. *J. Mater. Chem.*
688 *B Mater. Biol. Med. 11*, 2684–2692.
689 (48) Chen, D.-H., Baker, M. L., Hryc, C. F., DiMaio, F., Jakana, J., Wu, W., Dougherty,
690 M., Haase-Pettingell, C., Schmid, M. F., Jiang, W., Baker, D., King, J. A., and Chiu, W.
691 (2011) Structural basis for scaffolding-mediated assembly and maturation of a dsDNA
692 virus. *Proc Natl Acad Sci USA 108*, 1355–1360.
693 (49) Medina, E. M., Andrews, B. T., Nakatani, E., and Catalano, C. E. (2011) The
694 bacteriophage lambda gpNu3 scaffolding protein is an intrinsically disordered and
695 biologically functional procapsid assembly catalyst. *J. Mol. Biol. 412*, 723–736.
696 (50) Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M., and Eaton, W. A. (2005)
697 Polyproline and the “spectroscopic ruler” revisited with single-molecule fluorescence.
698 *Proc Natl Acad Sci USA 102*, 2754–2759.
699 (51) Cai, F., Bernstein, S. L., Wilson, S. C., and Kerfeld, C. A. (2016) Production and
700 Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins.
701 *Plant Physiol. 170*, 1868–1877.
702 (52) Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M., and Kerfeld, C.
703 A. (2019) Structure of a Synthetic β-Carboxysome Shell. *Plant Physiol. 181*, 1050–
704 1058.
705 (53) Li, S., Matoz-Fernandez, D. A., and Olvera de la Cruz, M. (2021) Effect of
706 mechanical properties on multicomponent shell patterning. *ACS Nano 15*, 14804–
707 14812.
708 (54) Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010)

709 Structural basis of the oxidative activation of the carboxysomal gamma-carbonic
710 anhydrase, CcmM. *Proc Natl Acad Sci USA 107*, 2455–2460.
711 (55) Heinhorst, S., Williams, E. B., Cai, F., Murin, C. D., Shively, J. M., and Cannon, G.
712 C. (2006) Characterization of the carboxysomal carbonic anhydrase CsoSCA from
713 Halothiobacillus neapolitanus. *J. Bacteriol. 188*, 8087–8094.
714 (56) Young, E. J., Burton, R., Mahalik, J. P., Sumpter, B. G., Fuentes-Cabrera, M.,
715 Kerfeld, C. A., and Ducat, D. C. (2017) Engineering the bacterial microcompartment
716 domain for molecular scaffolding applications. *Front. Microbiol. 8*, 1441.
717 (57) Mallette, E., and Kimber, M. S. (2017) A complete structural inventory of the
718 mycobacterial microcompartment shell proteins constrains models of global architecture
719 and transport. *J. Biol. Chem. 292*, 1197–1210.
720 (58) Ochoa, J. M., and Yeates, T. O. (2021) Recent structural insights into bacterial
721 microcompartment shells. *Curr. Opin. Microbiol. 62*, 51–60